Chapter 4
Bandwidth-Efficient Modulations with More Envelope Fluctuation

Thus far in our discussions, we have focused on constant or quasi-constant envelope modulations many of which, by virtue of their inherent memory, required a trellis decoder (as implemented by the VA [1]) for optimum reception. In theory, the VA can start producing a ML estimate of the transmitted signal only after observing the channel output corresponding to the entire transmitted signal, resulting in an infinite decoding delay. By decoding delay, we mean the amount of time (typically measured in number of bits) after which one begins to decode. Algorithms such as the truncated VA [2] can be used to reduce the decoding delay, but, in general, these lead to suboptimum receiver structures.

In certain applications, achieving a finite and small decoding delay is desirable. The natural question to ask is whether the requirement for finite decoding delay imposes constraints on the modulation/demodulation scheme that would reduce its optimality from a power and bandwidth-efficiency standpoint. Furthermore, to what extent would these constraints compromise the constant envelope nature of the transmitted signal set?

The ultimate goal would be to understand the possible trade-offs among minimum Euclidean distance (or, more generally, distance profile), bandwidth (or, more generally, PSD) and decoding delay. Such a goal is beyond the scope of this monograph. Instead, we consider here a reduced goal that investigates the above trade-offs for a particular structure derived from a generalization of that which implements MSK. The seeds for this investigation were planted in a paper presented at the 1997 International Symposium on Information Theory [3], in which Li and Rimoldi proposed a particular transmitter structure [the combination of

an encoder of memory, ν, and a waveform mapper—see Fig. 4-1(a)][1] for TCMs that, under certain constraints placed on the *differences* of the transmitted waveforms, guaranteed optimum decoding (using a conventional trellis decoder) with a finite (ν-bit duration) delay. Specifically, the encoder was simply a tapped delay line whose ν taps together with the input bit were mapped into a set of $M = 2^{\nu+1}$ waveforms (signals) of one-bit duration in accordance with a BCD relationship. That is, if $U_n \in 0, 1$ denotes the nth input bit and $U_{n-1}, U_{n-2}, \cdots, U_{n-\nu}$ the previous ν bits (the state of the encoder), then the signal transmitted in the interval $nT_b \leq t \leq (n+1)\,T_b$ would be $s_i\,(t)$, where the index, i, is defined in terms of these bits by $i = U_n \times 2^{\nu} + U_{n-1} \times 2^{\nu-1} + \cdots + U_{n-\nu-1} \times 2^1 + U_{n-\nu} \times 2^0$. It was also shown in Ref. 3 that, in addition to the constraints placed on the waveform differences, it was possible to further constrain the signals so as to

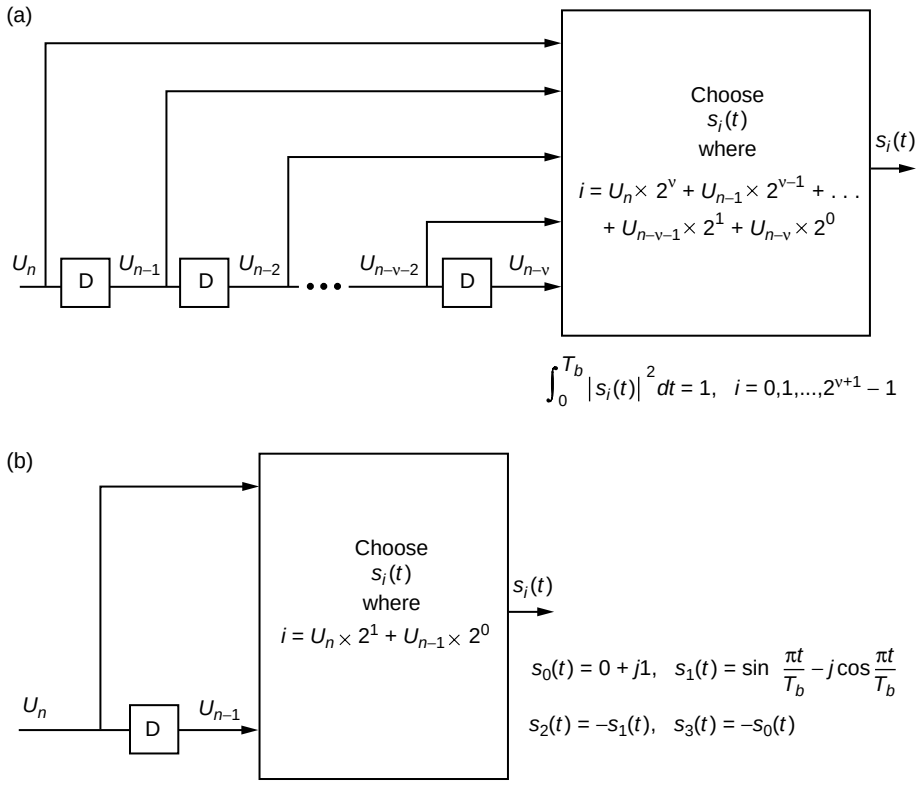

Fig. 4-1: (a) A trellis-coded modulation complex baseband transmitter and (b) the special case of "MSK" (ν = 1).

[1] Note the synergy of this structure with the TCM representation of FQPSK illustrated in Fig. 3-12.

maximize the value of the minimum squared Euclidean distance taken over all pairs of error event paths, namely, $d^2_{\min} = 2$. Such a maximum value of $d^2_{\min}$, which corresponds to a number of binary modulations such as BPSK and MSK, indicates that the receiver is providing optimum reception from a power conservation standpoint. Finally, in the presence of all of the above constraints, Li and Rimoldi [3] showed that it is possible to further optimize the system by selecting a set of waveforms that minimize the bandwidth-bit time product, BT_b.

In this chapter, we investigate an alternative (simpler) representation of the transmitter configuration suggested in Ref. 3 that consists of nothing more than a single filter (with complex impulse response) whose input is the ± 1 equivalent of the input data bits, namely, $\bar{U}_n = 1 - 2U_n$ for all n. This representation is arrived at by viewing the transmitted signal as a random pulse train with a pulse shape that extends beyond a single bit interval, i.e., one that contributes intersymbol interference (ISI) to its neighbors. As we shall see, such a pulse shape of duration $(\nu + 1) T_b$ can be constructed by designing its $\nu + 1$ partitions of duration T_b s in terms of the waveform differences that are outputted from Li and Rimoldi's transmitter. Such an ISI-based transmitter representation has the advantage that the PSD, and hence, the bandwidth are readily evaluated using known results for uncoded, random binary complex pulse trains. It also allows applying the insight provided in Forney's classic paper [4] on the VA, in particular, the discussion regarding the use of this algorithm to combat ISI.

One of the requirements placed on the set of possible transmitted waveforms $s_i(t), i = 0, 1, \cdots, M$ in Ref. 3 is that they all have equal energy.[2] Following consideration of the alternative representation described above, we discuss the impact of relaxing the equal energy restriction on the power efficiency of the modulation scheme in its ability to achieve the largest value of $d^2_{\min}$. In particular, we propose an additional set of constraints (now on the differences of the *energies* of the signals) that must be satisfied to achieve the same finite decoding delay, using again the optimum sequence receiver, and then demonstrate that such a set of constraints results in a signal design with a maximum value of $d^2_{\min}$ less than two. Allowing the signals to have unequal energy, however, suggests the possibility of additional flexibility in the design of these signals in order to achieve the best bandwidth efficiency. Thus, the reduction in $d^2_{\min}$ caused by the unequal energy requirement can possibly trade off against an additional reduction in signal bandwidth. Additional consideration of this notion warrants investigation.

---

[2] Note that the assumption of equal energy does not imply constant envelope, as was the case for the CPMs studied in Ref. 5, which served as the motivation for the work leading up to the results in Ref. 3. Nevertheless, the envelope fluctuation of the resulting signal designs will be small when compared with Nyquist designs of comparable bandwidth efficiencies, to be discussed later on.

4.1 Bandwidth-Efficient TCM with Prescribed Decoding Delay—Equal Signal Energies

4.1.1 ISI-Based Transmitter Implementation

The decomposition of a memory modulation into a cascade of an encoder and a memoryless modulator was first applied to CPM by Rimoldi [5]. In particular, for MSK (see Sec. 2.8.1.5*b* of this monograph), the transmitter obtained is illustrated in Fig. 2-18. Comparing Fig. 2-18 with the special case of Fig. 4-1(a), corresponding to $\nu = 1$ and illustrated in Fig. 4-1(b), we note that in the former, the state is represented by the differentially encoded version of the current input bit $V_n = U_n \oplus V_{n-1}$ whereas, in the latter, it would be just the previous input bit, U_{n-1} itself. Furthermore, because of the differential encoding associated with the state in Fig. 2-18, a differential decoder would be required in the receiver following the trellis decoder, which would result in a small loss in BEP performance. We have previously shown in Sec. 2.8.1.3 that precoding true MSK with a differential decoder at the transmitter results in a modulation that is equivalent (spectral and power efficiently) to MSK but without the need for differential decoding at the receiver. It is such precoded MSK that is implemented by the simpler configuration of Fig. 4-1(b) and denoted by the quotation marks around MSK in the caption. In what follows, when referring to MSK in the context of Fig. 4-1(b) or its equivalents, we shall assume that precoded MSK is implied.

Consider an uncoded random binary (± 1) sequence, $\{\bar{U}_n\}$, that generates a random pulse train

$$s'(t) = \sum_{n=-\infty}^{\infty} \bar{U}_n p(t - nT_b) \tag{4.1-1}$$

where $p(t) \triangleq p_R(t) + jp_I(t)$ is a complex pulse shape defined on the interval $0 \leq t \leq (\nu+1)T_b$. Consider partitioning $p(t)$ into $\nu+1$ adjoint pieces corresponding to its one-bit interval sections. That is, we define the set of T_b-s duration waveforms

$$p_k(t) \triangleq p_{Rk}(t) + jp_{Ik}(t) = \begin{cases} p(t+kT), & 0 \leq t \leq T_b \\ 0, & \text{otherwise} \end{cases}, \quad k = 0, 1, 2, \cdots, \nu \tag{4.1-2}$$

From (4.1-1), in any T_b-s interval, e.g., the nth, the signal $s'(t)$ will be described by one of $M = 2^{\nu+1}$ complex waveforms, i.e., $s'_k(t - nT_b), k = 0, 1, 2, \cdots, 2^{\nu+1} - 1$, which are expressed in terms of $p(t)$ and the data sequence, $\{\bar{U}_n\}$, by

$$s'_k(t-nT_b) = \bar{U}_n p_0(t-nT_b) + \bar{U}_{n-1} p_1(t-nT_b) + \cdots + \bar{U}_{n-\nu} p_\nu(t-nT_b),$$

$$k = 0, 1, 2, \cdots, 2^{\nu+1} - 1 \tag{4.1-3}$$

where the index, k, is the equivalent (0,1) bit sequence $\{U_n, U_{n-1}, \cdots, U_{n-\nu}\}$ expressed in BCD form. As an example, the set of waveforms for memory $\nu = 2$ is given below:

$$\left.\begin{aligned}
s'_0(t-nT_b) &= p_0(t-nT_b) + p_1(t-nT_b) + p_2(t-nT_b) \\
s'_1(t-nT_b) &= p_0(t-nT_b) + p_1(t-nT_b) - p_2(t-nT_b) \\
s'_2(t-nT_b) &= p_0(t-nT_b) - p_1(t-nT_b) + p_2(t-nT_b) \\
s'_3(t-nT_b) &= p_0(t-nT_b) - p_1(t-nT_b) - p_2(t-nT_b) \\
s'_4(t-nT_b) &= -p_0(t-nT_b) + p_1(t-nT_b) + p_2(t-nT_b) \\
s'_5(t-nT_b) &= -p_0(t-nT_b) + p_1(t-nT_b) - p_2(t-nT_b) \\
s'_6(t-nT_b) &= -p_0(t-nT_b) - p_1(t-nT_b) + p_2(t-nT_b) \\
s'_7(t-nT_b) &= -p_0(t-nT_b) - p_1(t-nT_b) - p_2(t-nT_b)
\end{aligned}\right\} \tag{4.1-4}$$

We note from (4.1-4) that, because of the BCD construction, the following properties hold for the signal differences:

$$s'_0(t) - s'_1(t) = s'_2(t) - s'_3(t) = s'_4(t) - s'_5(t) = s'_6(t) - s'_7(t) = 2p_2(t) \tag{4.1-5a}$$

$$s'_0(t) - s'_2(t) = s'_4(t) - s'_6(t) = 2p_1(t) \tag{4.1-5b}$$

Also, an equivalent (at least insofar as the first equality is concerned) condition to (4.1-5b) is

$$s'_0(t) - s'_4(t) = s'_2(t) - s'_6(t) = 2p_0(t) \tag{4.1-5c}$$

In the more generic case for arbitrary ν, the conditions corresponding to (4.1-5a) and (4.1-5b) would be summarized as:

$$s'_0(t) - s'_{2^m}(t) = s'_{2^{m+1}l}(t) - s'_{2^{m+1}l+2^m}(t) = 2p_{\nu-m}(t),$$

$$m = 0, 1, 2, \cdots, \nu - 1, \quad l = 1, 2, \cdots, 2^{\nu-m} - 1 \tag{4.1-6}$$

and, furthermore, the generalization of (4.1-5c) becomes

$$s'_0(t) - s'_{2^\nu}(t) = s'_{2^{\nu-1}}(t) - s'_{2^\nu+2^{\nu-1}}(t) = 2p_0(t) \tag{4.1-7}$$

Associating the $2^{\nu+1}$ signals $\{s'_k(t)\}$ expressed as in (4.1-3) with the assumed equal energy, $\{s_k(t)\}$, derived from the implementation in Fig. 4-1(a), we see that the conditions on the signal differences of $s'_i(t)$ given in (4.1-6) are precisely those of Theorem I in Ref. 3, which guarantees a finite decoding delay of ν bits using an optimum trellis-coded receiver.[3] Therefore, since $p(t)$ is entirely specified by its adjoint T_b-s sections, T_b, it would appear that the transmitter of Fig. 4-1(a) can be equivalently implemented [see Fig. 4-2(a)] by passing the input ± 1 data sequence, $\{\bar{U}_n\}$ (modeled as a random impulse train), through a filter with complex impulse response

$$\left.\begin{aligned} p(t) &= \sum_{i=0}^{\nu} p_i(t - iT_b) \\ p_i(t) &= \frac{1}{2}\left[s'_0(t) - s'_{2^{\nu-i}}(t)\right] \end{aligned}\right\} \tag{4.1-8}$$

or equivalently [see Fig. 4-2(b)], the real and imaginary parts of the base-band signal (to be modulated onto quadrature carriers for transmission over the

[3] Li and Rimoldi also note that these conditions guarantee that the Euclidean distance between any pair of paths in the trellis decoder diverging at time n and remerging at time $n + \nu + 1$ is the same. Furthermore, the number of correlators (matched filters) needed to implement the optimum (MLSE) receiver will now vary *linearly* with memory, i.e., $\nu + 1$, as opposed to *exponentially* with memory, i.e., $2^{\nu+1}$, which is the case when no constraints are imposed on the decoding delay.

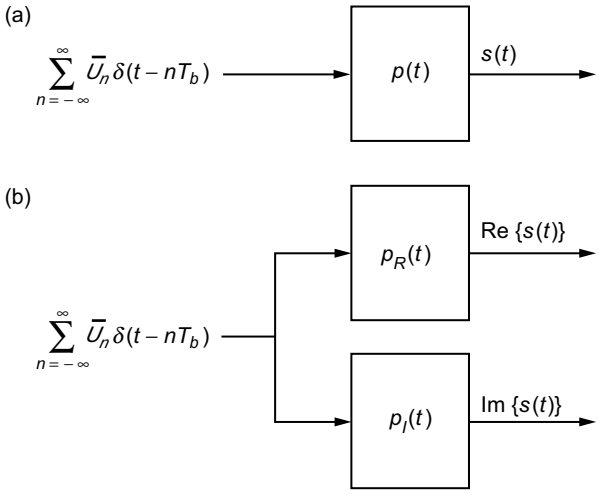

Fig. 4-2: (a) Complex baseband transmitter for MSK equivalent to Fig. 4-1(b) and (b) an I-Q baseband transmitter for MSK equivalent to Fig. 4-1(b).

channel) can be obtained by passing the common input ± 1 data sequence, $\{\bar{U}_n\}$, through a pair of filters with respective impulse responses

$$\left.\begin{aligned} p_{Ri}(t) &= \frac{1}{2}\left[s'_{R0}(t) - s'_{R2^{\nu-i}}(t)\right] \\ p_{Ii}(t) &= \frac{1}{2}\left[s'_{I0}(t) - s'_{I2^{\nu-i}}(t)\right] \end{aligned}\right\} \tag{4.1-9}$$

Unfortunately, the implementation in Fig. 4-2(a) is not always equivalent to that in Fig. 4-1(a), but as we shall see momentarily, for the case of most practical interest, i.e., a signal set $\{s_k(t)\}$ with maximum minimum Euclidean distance between its members, the equivalence between the two implementations is guaranteed, i.e., $\{s'_k(t)\}$ and $\{s_k(t)\}$ are identical. Before showing this, we note that even though $p_R(t)$ and $p_I(t)$ are constructed from the real and imaginary components of a set of equal energy complex signals, $\{s'_k(t), k = 0, 1, 2, \cdots, 2^{\nu+1} - 1\}$, they themselves do not necessarily have equal energy. We shall see that this is true, even for the simple case of MSK.

Note that because of the symmetry of the BCD mapping, the signals in the memory two example of (4.1-4) also satisfy the conditions

$$\left.\begin{aligned} s'_0(t) &= -s'_7(t) \\ s'_1(t) &= -s'_6(t) \\ s'_2(t) &= -s'_5(t) \\ s'_3(t) &= -s'_4(t) \end{aligned}\right\} \tag{4.1-10}$$

which, in the case of arbitrary memory, ν, would become

$$s'_m(t) = -s'_{2^{\nu+1}-1-m}(t), \quad m = 0, 1, \cdots, 2^\nu - 1 \tag{4.1-11}$$

The conditions of (4.1-11), which correspond to an antipodal signaling set, are precisely those given in Ref. 3. They achieve the maximum value of minimum-squared Euclidean distance, namely, $d^2_{\min} = 2$. Thus, the implementation of Fig. 4-2(a) not only achieves finite decoding delay but also automatically achieves the optimum performance from the standpoint of power efficiency. This result should not be surprising in view of the findings in Ref. 4, which indicate that an MLSE-form of receiver such as the trellis decoder can completely remove ISI and thereby achieve the performance of a zero-ISI (full-response) system. However, since the implementation in Fig. 4-1(a) can produce a set of signals, $\{s_k(t)\}$, that satisfies the difference properties needed for finite decoding delay without requiring them to have maximum minimum Euclidean distance, then the two implementations will be equivalent, i.e., $\{s_k(t)\} = \{s'_k(t)\}$ only when this additional requirement is imposed. A formal proof of this equivalence is presented in Ref. 3. In what follows, we consider only the important practical case of antipodal signal sets and, as such, drop the prime notation on the signals derived from $p(t)$.

What remains is to consider the bandwidth efficiency of signals designed according to the constraints of (4.1-6), (4.1-7), and (4.1-11). This is where the ISI-based representation of Fig. 4-2(a) helps considerably, since the evaluation of the PSD of the transmitted signal can be trivially accomplished using well-known relations [6] for random pulse trains. This is considered in the next section.

4.1.2 Evaluation of the Power Spectral Density

In this section, we compute the PSD of a random complex pulse train, e.g., that in (4.1-1), modulated onto quadrature carriers. That is, if the transmitted bandpass signal is given by[4]

$$\begin{aligned}\tilde{s}(t) &= \mathrm{Re}\left\{s(t)\, e^{j2\pi f_m t}\right\} \\ &= \left(\sum_{n=-\infty}^{\infty} \bar{U}_n p_R(t - nT_b)\right) \cos 2\pi f_m t \\ &\quad - \left(\sum_{n=-\infty}^{\infty} \bar{U}_n p_I(t - nT_b)\right) \sin 2\pi f_m t \end{aligned} \tag{4.1-12}$$

then it is straightforward to show using an extension of the methods in Chap. 2 of Ref. 6 that the PSD of $\tilde{s}(t)$ is given by

$$\begin{aligned} S(f) &= \frac{1}{4T_b} \left|P_R(f - f_m) + jP_I(f - f_m)\right|^2 \\ &\quad + \frac{1}{4T_b} \left|P_R(f + f_m) - jP_I(f + f_m)\right|^2 \\ &= S_u(f) + S_l(f) \end{aligned} \tag{4.1-13}$$

where

$$\left.\begin{aligned} P_R(f) &\triangleq \mathcal{F}\{p_R(t)\} \\ P_R(f) &\triangleq \mathcal{F}\{p_I(t)\} \end{aligned}\right\} \tag{4.1-14}$$

are the Fourier transforms of the real and imaginary pulse shapes which, in general, are complex functions of f, and the u and l subscripts denote upper and lower sideband, respectively. Note that the signal in (4.1-12) differs from the usual QPSK-type of signal in that here, the same data sequence is passed

[4] We use the notation f_m for the actual modulating frequency of the quadrature carriers to distinguish it from the carrier frequency around which the PSD is symmetric, which will be denoted by f_c. More about this shortly.

through both the I and Q filters whereas for QPSK, the two sequences passing through these filters would be different and independent of one another. As such, the PSD in (4.1-13) cannot, in general, be written in the form [6, Eq. (2.131)]

$$S(f) = \frac{1}{4}G(f - f_c) + \frac{1}{4}G(f + f_c) \tag{4.1-15}$$

where $G(f)$ is the equivalent baseband (symmetrical around $f = 0$) PSD and is a real function of f, and f_c is some arbitrary carrier frequency.[5]

To demonstrate the above point, consider the specific case of MSK ($\nu = 1$), for which the four complex signals are given by[6]

$$\left.\begin{aligned} s_0(t) &= 0 + j1 \\ s_1(t) &= \sin\frac{\pi t}{T_b} - j\cos\frac{\pi t}{T_b} = s_0^*(t)\, e^{j\frac{\pi t}{T_b}} \\ s_2(t) &= -s_1(t) \\ s_3(t) &= -s_0(t) \end{aligned}\right\} \tag{4.1-16}$$

In terms of the ISI-based representation, we obtain from (4.1-8) that

$$\left.\begin{aligned} p_0(t) &= \frac{1}{2}\sin\frac{\pi t}{T_b} + j\frac{1}{2}\left[1 - \cos\frac{\pi t}{T_b}\right] \\ p_1(t) &= -\frac{1}{2}\sin\frac{\pi t}{T_b} + j\frac{1}{2}\left[1 + \cos\frac{\pi t}{T_b}\right] \end{aligned}\right\} \tag{4.1-17}$$

---

[5] What is meant by an "equivalent baseband PSD" is a PSD around zero frequency that is *identical* to the upper or lower sideband of the bandpass PSD, frequency-shifted to the origin. While it is always possible to express (4.1-13) in the form $S(f) = (1/4)G_u(f - f_c) + (1/4)G_l(f + f_c)$ where $G_u(f) = G_l(-f)$, in general, there is no guarantee that $G_u(f)$ [or equivalently, $G_l(f)$] has symmetry about the origin, or for that matter, about any frequency f_c. Stated another way, while demodulating the bandpass signal with a carrier at some frequency f_c (not necessarily equal to the modulating frequency f_m) will always produce a symmetric PSD around the origin, the resulting baseband PSD will, in general, be a combination (sum) of the aliased upper and lower sidebands, and may or may not appear as a simple frequency translation of either of these sidebands.

[6] Note that for the Rimoldi decomposition of MSK illustrated in Fig. 2-18, the signals satisfy the condition $s_0(t) - s_1(t) = -(s_2(t) - s_3(t))$ rather than $s_0(t) - s_1(t) = s_2(t) - s_3(t)$, as required by (4.1-5a), for the signals of (4.1-16) corresponding to precoded MSK.

Thus, using (4.1-17) to define the complex pulse shape of (4.1-8), we obtain

$$p(t) = \frac{1}{2}\sin\frac{\pi t}{T_b} + j\frac{1}{2}\left[1 - \cos\frac{\pi t}{T_b}\right], \quad 0 \le t \le 2T_b \tag{4.1-18}$$

That is, an appropriate implementation for MSK that guarantees a decoding delay of one bit is that of Fig. 4-2(b), with I and Q filters having impulse responses

$$\left.\begin{aligned} p_R(t) &= \frac{1}{2}\sin\frac{\pi t}{T_b}, \quad 0 \le t \le 2T_b \\ p_I(t) &= \frac{1}{2}\left[1 - \cos\frac{\pi t}{T_b}\right], \quad 0 \le t \le 2T_b \end{aligned}\right\} \tag{4.1-19}$$

Taking the Fourier transforms of $p_R(t)$ and $p_I(t)$ of (4.1-8) and using these in (4.1-13), we arrive at the following result for the bandpass PSD:

$$\begin{aligned} S(f) &= \frac{T_b}{4}\frac{\sin^2 2\pi(f - f_m)T_b}{\pi^2}\left[\frac{1}{1 - 2(f - f_m)T_b} + \frac{1}{2(f - f_m)T_b}\right]^2 \\ &\quad + \frac{T_b}{4}\frac{\sin^2 2\pi(f + f_m)T_b}{\pi^2}\left[\frac{1}{1 + 2(f + f_m)T_b} - \frac{1}{2(f + f_m)T_b}\right]^2 \\ &= S_u(f) + S_l(f) \end{aligned} \tag{4.1-20}$$

Note that while $S(f)$ is an even function of f (as it should be for a real signal), its upper and lower sidebands, $S_u(f)$ and $S_l(f)$, are not symmetric around f_m and $-f_m$, respectively. However, there does exist a frequency, $f_c \neq f_m$, around which the upper sideband (and similarly for the lower sideband) is symmetric. To understand why this is so, we remind the reader that according to Rimoldi's decomposition [5], the modulation frequency chosen for the quadrature carriers should be shifted from the carrier frequency f_c, around which the bandpass spectrum is to be symmetric by an amount equal to $1/4T_b$, i.e., $f_m = f_c - 1/4T_b$. This stems from the fact that the specification of the signals as in (4.1-16) results in a tilted trellis where the phase tilt is equal to $\pi/2$ rad. (Note that a frequency shift of $\Delta f = 1/4T_b$ is equal to a phase shift $2\pi\Delta fT = \pi/2$). To demonstrate that this is indeed the case, we evaluate the PSD of MSK, using (4.1-20) with the shifted value of modulating frequency, $f_m = f_c - 1/4T_b$. When this is done, the result in (4.1-15) is obtained with

$$G(f) = \frac{16T_b}{\pi^2} \frac{\cos^2 2\pi f T_b}{\left(1 - 16f^2 T_b^2\right)^2} \tag{4.1-21}$$

which corresponds (except for a normalization factor) to the well-known PSD of MSK [6, Eq. (2.148)].

The question that comes about now is: For arbitrary memory, ν, and a baseband signal design satisfying (4.1-6), (4.1-7), and (4.1-11), is it possible to find a modulating frequency, f_m, that will produce a symmetric bandpass PSD around some other carrier frequency, f_c? If not, then one cannot find an equivalent baseband PSD, and, hence, the bandwidth (whatever measure is used) of the signal must be determined from the RF waveform.

4.1.2.1 The Memory One Case. To shed some light on the answer to the above question, we consider the simplest case of unit memory, where the complex pulse shape of (4.1-8) is given by

$$\begin{aligned} p(t) &= \frac{1}{2}\left[s_0(t) - s_2(t) + s_0(t - T_b) - s_1(t - T_b)\right] \\ &= \frac{1}{2}\left[s_0(t) + s_0(t - T_b) + s_1(t) + s_2(t - T_b)\right], \quad 0 \le t \le 2T_b \end{aligned} \tag{4.1-22}$$

where, in accordance with (4.1-11), we have used the fact that $s_1(t) = -s_2(t)$ in order to achieve $d_{\min}^2 = 2$. The Fourier transform of $p(t)$ in (4.1-22) is given by

$$\begin{aligned} P(f) = \frac{1}{2}\Bigg[&\int_0^{T_b} s_0(t)\left(1 + e^{-j2\pi f T_b}\right) e^{-j2\pi f t} dt \\ &+ \int_0^{T_b} s_1(t)\, e^{-j2\pi f t} dt + e^{-j2\pi f T_b} \int_0^{T_b} s_2(t)\, e^{-j2\pi f t} dt\Bigg] \end{aligned} \tag{4.1-23}$$

Since from (4.1-13), the upper spectral sideband is $S_u(f) = (1/4T_b)\left|P(f - f_m)\right|^2$, then in order for this to be symmetric around f_c, we must have

$$\left|P(f_c + f - f_m)\right|^2 = \left|P(f_c - f - f_m)\right|^2 \tag{4.1-24}$$

or letting $f_s \triangleq f_c - f_m$ denote the separation between the actual modulation frequency and the bandpass frequency around which symmetry is desired, $s_0(t)$ and $s_1(t)$ must be chosen to satisfy

$$|P(f_s+f)|^2 = |P(f_s-f)|^2 \tag{4.1-25a}$$

or equivalently

$$|P(f_s+f)|^2 = |P^*(f_s-f)|^2 \tag{4.1-25b}$$

for some f_s. In terms of (4.1-23), the spectral equality in (4.1-25b) requires that we have

$$\begin{aligned}
&\left| \int_0^{T_b} \left(s_0(t)\, e^{-j2\pi f_s t}\right) e^{-j2\pi f t} dt + e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} \left(s_0(t)\, e^{-j2\pi f_s t}\right) e^{-j2\pi f t} dt \right. \\
&\left. + \int_0^{T_b} \left(s_1(t)\, e^{-j2\pi f_s t}\right) e^{-j2\pi f t} dt + e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} \left(s_2(t)\, e^{-j2\pi f_s t}\right) e^{-j2\pi f t} dt \right|^2 \\
&= \left| \int_0^{T_b} \left(s_0^*(t)\, e^{j2\pi f_s t}\right) e^{-j2\pi f t} dt + e^{j2\pi(f_s-f)T_b} \int_0^{T_b} \left(s_0^*(t)\, e^{j2\pi f_s t}\right) e^{-j2\pi f t} dt \right. \\
&\left. + \int_0^{T_b} \left(s_1^*(t)\, e^{j2\pi f_s t}\right) e^{-j2\pi f t} dt + e^{j2\pi(f_s-f)T_b} \int_0^{T_b} \left(s_2^*(t)\, e^{j2\pi f_s t}\right) e^{-j2\pi f t} dt \right|^2
\end{aligned} \tag{4.1-26}$$

Sufficient conditions on the signals, $\{s_i(t)\}$, for (4.1-26) to be satisfied are

$$\left.\begin{aligned} s_1(t) &= s_0^*(t)\, e^{j4\pi f_s t} \\ s_2(t) &= e^{j4\pi f_s T_b} s_0^*(t)\, e^{j4\pi f_s t} \end{aligned}\right\} \tag{4.1-27}$$

However, since in arriving at (4.1-26), we have already assumed that $s_1(t) = -s_2(t)$, then (4.1-27) further requires that $f_s = 1/4T_b$, from which we obtain the complete signal set

$$\left.\begin{aligned} s_1(t) &= s_0^*(t)\, e^{j\pi t/T_b} \\ s_2(t) &= -s_0^*(t)\, e^{j\pi t/T_b} \\ s_3(t) &= -s_0(t) \end{aligned}\right\} \tag{4.1-28}$$

Note that for memory one it is only necessary to specify $s_0(t)$ in order to arrive at the complete signal set. Also, the signal set of (4.1-28) satisfies the finite decoding delay condition of Ref. 3, namely, $s_0(t) - s_1(t) = s_2(t) - s_3(t)$.

The equivalent lowpass PSD is obtained by first using $s_1(t) = -s_2(t)$ in (4.1-23), resulting in

$$P(f) = \frac{1}{2}\Big[S_0(f) + S_1(f) + e^{-j2\pi f T_b}\big(S_0(f) - S_1(f)\big)\Big] \tag{4.1-29}$$

from which one immediately gets

$$\frac{1}{T_b}|P(f)|^2 =$$

$$\frac{1}{2T_b}\left[|S_0(f)|^2 + |S_1(f)|^2 + \mathrm{Re}\Big\{\big(S_0^*(f) + S_1^*(f)\big)\big(S_0(f) - S_1(f)\big)e^{-j2\pi f T_b}\Big\}\right] \tag{4.1-30}$$

In (4.1-29) and (4.1-30), $S_i(f)$ denotes the Fourier transform of $s_i(t)$. Using the first symmetry condition of (4.1-28) in (4.1-30) gives the desired equivalent lowpass PSD, namely,

$$\begin{aligned} \frac{1}{T_b}\left|P\left(f+\frac{1}{4T_b}\right)\right|^2 &= \left|S_0\left(f+\frac{1}{4T_b}\right)\right|^2 [1-\sin 2\pi f T_b] \\ &\quad + \left|S_0\left(-f+\frac{1}{4T_b}\right)\right|^2 [1+\sin 2\pi f T_b] \\ &\quad + 2\left[\mathrm{Re}\left\{S_0\left(f+\frac{1}{4T_b}\right)\right\}\mathrm{Im}\left\{S_0\left(-f+\frac{1}{4T_b}\right)\right\}\right. \\ &\quad \left. + \mathrm{Re}\left\{S_0\left(-f+\frac{1}{4T_b}\right)\right\}\mathrm{Im}\left\{S_0\left(f+\frac{1}{4T_b}\right)\right\}\right]\cos 2\pi f T_b \end{aligned} \tag{4.1-31}$$

which is clearly an even function of frequency.

Although (4.1-28) is satisfied by the MSK signals of (4.1-16) as should be the case, this condition applies in a more general context since it does not explicitly specify $s_0(t)$ but rather only the *relation between* $s_0(t)$ and $s_1(t)$. This should not be surprising since it has been shown in the past that there exists an entire class of MSK-type signals (referred to in Ref. 7 as generalized MSK) which happen to also be constant envelope (in addition to being equal energy) and achieve $d^2_{\min} = 2$ as well as a decoding delay of one bit interval. In particular, the class of binary full-response CPM signals with modulation index $h = 1/2$ and equivalent phase pulse $q(t)$, which satisfies the conditions of (2.8-5), is appropriate, an example of which is Amoroso's SFSK [8] for which $q(t)$ is given by (2.8-9).

4.1.2.2 The Memory Two Case. For memory two, the pulse shape is given by

$$p(t) = \frac{1}{2}\left[s_0(t) - s_4(t) + s_0(t-T_b) - s_2(t-T_b) + s_0(t-2T_b) - s_1(t-2T_b)\right]$$

$$= \frac{1}{2}\left[s_0(t) + s_0(t-T_b) + s_0(t-2T_b) + s_3(t) - s_2(t-T_b) - s_1(t-2T_b)\right],$$

$$0 \le t \le 3T_b \qquad (4.1\text{-}32)$$

with Fourier transform

$$P(f) = \frac{1}{2}\left[\left(1 + e^{-j2\pi fT_b} + e^{-j4\pi fT_b}\right)\int_0^{T_b} s_0(t)\, e^{-j2\pi ft}dt + \int_0^{T_b} s_3(t)\, e^{-j2\pi ft}dt\right.$$

$$\left. - e^{-j2\pi fT_b}\int_0^{T_b} s_2(t)\, e^{-j2\pi ft}dt - e^{-j4\pi fT_b}\int_0^{T_b} s_1(t)\, e^{-j2\pi ft}dt\right] \qquad (4.1\text{-}33)$$

Applying (4.1-33) to (4.1-25b) and letting $s_3(t) = s_2(t) - s_0(t) + s_1(t)$, in accordance with (4.1-5a), we obtain the bandpass spectral symmetry condition

$$\left| e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} \left(s_0(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt + e^{-j4\pi(f_s+f)T_b} \right.$$

$$\times \int_0^{T_b} \left(s_0(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt$$

$$+ \int_0^{T_b} \left(s_2(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt - e^{-j2\pi(f_s+f)T_b} \int_0^{T_b} \left(s_2(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt$$

$$\left. + \int_0^{T_b} \left(s_1(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt - e^{-j4\pi(f_s+f)T_b} \int_0^{T_b} \left(s_1(t)\, e^{-j2\pi f_s t} \right) e^{-j2\pi f t} dt \right|^2$$

$$= \left| e^{j2\pi(f_s-f)T_b} \int_0^{T_b} \left(s_0^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt + e^{j4\pi(f_s-f)T_b} \right.$$

$$\times \int_0^{T_b} \left(s_0^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt$$

$$+ \int_0^{T_b} \left(s_2^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt - e^{j2\pi(f_s-f)T_b} \int_0^{T_b} \left(s_2^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt$$

$$\left. + \int_0^{T_b} \left(s_1^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt - e^{j4\pi(f_s-f)T_b} \int_0^{T_b} \left(s_1^*(t)\, e^{j2\pi f_s t} \right) e^{-j2\pi f t} dt \right|^2 \tag{4.1-34}$$

Analogous with (4.1-27), satisfying (4.1-34) implies the set of conditions

$$s_1(t) + s_2(t) = \left(s_1^*(t) + s_2^*(t) \right) e^{j4\pi f_s t} \tag{4.1-35a}$$

$$s_0(t) - s_2(t) = e^{j4\pi f_s T_b} \left(s_0^*(t) - s_2^*(t) \right) e^{j4\pi f_s t} \tag{4.1-35b}$$

$$s_0(t) - s_1(t) = e^{j8\pi f_s T_b} \left(s_0^*(t) - s_1^*(t) \right) e^{j4\pi f_s t} \tag{4.1-35c}$$

Again letting $f_s = 1/4T_b$ and summing (4.1-35a), (4.1-35b), and (4.1-35c) gives

$$s_1(t) + s_2(t) = \big(s_1^*(t) + s_2^*(t)\big)e^{j\pi t/T_b} \tag{4.1-36a}$$

$$s_0(t) = s_2^*(t)\, e^{j\pi t/T_b} \quad \Big(\text{or equivalently } s_2(t) = s_0^*(t)\, e^{j\pi t/T_b}\Big) \tag{4.1-36b}$$

$$s_0(t) - s_1(t) = \big(s_0^*(t) - s_1^*(t)\big)e^{j\pi t/T_b} \tag{4.1-36c}$$

Actually, (4.1-36c) is not an independent condition since it can be derived from (4.1-36a) and (4.1-36b). Thus, (4.1-36a) and (4.1-36b) are sufficient to determine the signal design.

Following along the lines of (4.1-29) and (4.1-30), the equivalent PSD of the memory two modulation may be found. In particular, the Fourier transform of the equivalent pulse shape in (4.1-8) is given as

$$P(f) =$$

$$\frac{1}{2}\Big[S_0(f) + S_3(f) + e^{-j2\pi fT_b}\big(S_0(f) - S_2(f)\big) + e^{-j4\pi fT_b}\big(S_0(f) - S_1(f)\big)\Big] \tag{4.1-37}$$

Using the additional relation, $S_3(f) = S_1(f) + S_2(f) - S_0(f)$, to achieve finite decoding delay, one immediately gets the desired equivalent lowpass PSD as

$$\frac{1}{T_b}\left|P\left(f + \frac{1}{4T_b}\right)\right|^2 =$$

$$\frac{1}{4T_b}\left[\left|S_1\left(f + \frac{1}{4T_b}\right) + S_2\left(f + \frac{1}{4T_b}\right)\right|^2\right.$$

$$+ \left|S_0\left(f + \frac{1}{4T_b}\right) - S_2\left(f + \frac{1}{4T_b}\right)\right|^2 + \left|S_0\left(f + \frac{1}{4T_b}\right) - S_1\left(f + \frac{1}{4T_b}\right)\right|^2$$

$$+ 2\mathrm{Re}\left\{\left[S_1^*\left(f + \frac{1}{4T_b}\right) + S_2^*\left(f + \frac{1}{4T_b}\right)\right]\right.$$

$$\left.\left.\times\left[S_0\left(f + \frac{1}{4T_b}\right) - S_2\left(f + \frac{1}{4T_b}\right)\right]e^{-j2\pi(f+[1/4T_b])T_b}\right\}\right.$$

$$+ 2\mathrm{Re}\left\{\left[S_0^*\left(f+\frac{1}{4T_b}\right) - S_2^*\left(f+\frac{1}{4T_b}\right)\right]\right.$$

$$\left.\times\left[S_0\left(f+\frac{1}{4T_b}\right) - S_1\left(f+\frac{1}{4T_b}\right)\right]e^{-2\pi(f+[1/4T_b])T_b}\right\}$$

$$+ 2\mathrm{Re}\left\{\left[S_1^*\left(f+\frac{1}{4T_b}\right) + S_2^*\left(f+\frac{1}{4T_b}\right)\right]\right.$$

$$\left.\left.\times\left[S_0\left(f+\frac{1}{4T_b}\right) - S_1\left(f+\frac{1}{4T_b}\right)\right]e^{-4\pi(f+[1/4T_b])T_b}\right\}\right] \tag{4.1-38}$$

which, when (4.1-36) is used, can be shown to be an even function of frequency, as is necessary.

4.1.3 Optimizing the Bandwidth Efficiency

Having obtained expressions for the equivalent baseband PSD, it is now straightforward to use these to determine the sets of signals that satisfy all of the previous constraints and, in addition, maximize the power within a given bandwidth, B. In mathematical terms, we search for the set of signals that for a given value of B maximizes the fractional in-band power

$$\eta = \frac{\int_{-B/2}^{B/2} G(f)\,df}{\int_{-\infty}^{\infty} G(f)\,df}, \quad G(f) \triangleq \frac{1}{T_b}\left|P\left(f+\frac{1}{4T_b}\right)\right|^2 \tag{4.1-39}$$

subject to the unit power constraint

$$\frac{1}{T_b}\int_0^{T_b} |s_i(t)|^2\,dt = \frac{1}{T_b}\int_{-\infty}^{\infty} |S_i(f)|^2\,dt = 1, \quad i = 0, 1, 2, \cdots, M-1 \tag{4.1-40}$$

4.1.3.1 Memory One Case. For the case of $\nu = 1$, we observed that the entire signal set may be determined from the single complex signal, $s_0(t)$. Thus, optimizing bandwidth efficiency corresponds to substituting the PSD of (4.1-31) (which is entirely specified in terms of the Fourier transform of $s_0(t)$) into (4.1-39) and then maximizing η subject to (4.1-40). Such a procedure would result in an optimum $S_0(f)$ from whose inverse Fourier transform one could determine the optimum signal set. Since $S_0(f)$ exists, in general, over the entire doubly infinite frequency axis, it is perhaps simpler to approach the optimization in the

time domain, since $s_0(t)$ is indeed time limited to the interval $0 \leq t \leq T_b$. To do this, we need to first rewrite the PSD of (4.1-40) in terms of $s_0(t)$ rather than $S_0(f)$ and then perform the integrations on f required in (4.1-39). After considerable manipulation, and for simplicity of notation normalizing $T_b = 1$ (i.e., $BT_b = B$), it can be shown that

$$\int_{-B/2}^{B/2} G(f)\,df =$$

$$B\int_0^1\int_0^1 s_0(t)\,s_0^*(\tau)\,e^{-j(\pi/2)(t-\tau)}$$

$$\times\left[\text{sinc}\,\pi B(t-\tau) - j\frac{1}{2}\text{sinc}\,\pi B(t-\tau+1) + j\frac{1}{2}\text{sinc}\,\pi B(t-\tau-1)\right]dtd\tau$$

$$+\frac{1}{2}B\;\text{Im}\left\{\int_0^1\int_0^1 s_0(t)\,s_0(\tau)\,e^{-j(\pi/2)(t+\tau)}\right.$$

$$\left.\times\left[\text{sinc}\,\pi B(t-\tau+1) + \text{sinc}\,\pi B(t-\tau-1)\right]dtd\tau\right\} \tag{4.1-41}$$

where $\text{sinc}\,x \triangleq \sin x/x$. Furthermore, it is straightforward to show that

$$\int_{-\infty}^{\infty} G(f)\,df = 1 \tag{4.1-42}$$

and, thus, η is given directly by (4.1-41).

The maximization of (4.1-41) subject to the energy constraint of (4.1-40) has been carried out numerically, using the MATLAB$^{(\text{r})}$ (software application) optimization toolbox function "fminunc" (quasi-Newton method of convergence). In particular, for each value of B (BT_b if $T_b \neq 1$), the optimum complex signal, $s_0(t)$, [represented by N uniformly spaced samples in the interval $(0,1)$], is determined, from which the fractional out-of-band power, $1-\eta$, is calculated using (4.1-41) for η. Because of complexity issues involved in computing the optimum solution, the number of sample points, N, is limited to 64. Furthermore, since the Gaussian integration required to evaluate with high accuracy the double integral of (4.1-41) requires a much higher density of sample values (not necessarily uniformly spaced), then to allow for Fourier interpolation, we assume the signal

to be bandlimited[7] to the Nyquist rate, i.e., 32 ($32/T_b$ if $T_b \neq 1$). Because of this bandlimiting assumption, certain optimum signal waveforms (particularly those at small values of B) that exhibit a sharp discontinuity will have a ringing behavior. This ringing behavior can be minimized by additional interpolation (filtering) but has proven difficult to eliminate completely.

Figures 4-3(a) and 4-3(b) are 3-D plots of the optimum real and imaginary parts of $s_0(t)$ versus t as a function of B in the interval $0 \leq B \leq 3$. Figures 4-4(a)–(h) are a number of cuts of these 3-D plots taken at distinct values of B in the same range. For small values of B, we observe that the real part of $s_0(t)$ has sharp discontinuities at $t = 0$ and $t = 1$ and, thus, exhibits the ringing behavior alluded to above. As B increases, the sharpness of the discontinuity at the edges diminishes, and in the limit of large B, both the real and imaginary parts of $s_0(t)$ approach a sinusoid with unit period. Specifically, $s_0(t)$ tends toward the form $-\alpha_1 \sin 2\pi t + j(\beta_1 + \alpha_2 \cos 2\pi t)$, where $\alpha_1, \alpha_2, \beta_1$ are constants that also must satisfy the unit energy constraint, i.e., $\beta_1^2 + (1/2)\left(\alpha_1^2 + \alpha_2^2\right) = 1$. Figure 4-5 is the corresponding plot of optimum (minimum) fractional out-of-band power versus B. Also shown are corresponding results for MSK and SFSK modulations that can readily be found in Fig. 2.11 of Ref. 6.[8] We observe that by optimizing the signal set at each value of B without loss in $d_{\min}^2$ or finite decoding delay performance, we are able to obtain a significant improvement in bandwidth efficiency. The quantitative amount of this improvement is given in Table 4-1 for the 99 percent and 99.9 percent bandwidths corresponding respectively to the -20 dB and -30 dB out-of-band power levels.

Before concluding this section, we note that the maximization of (4.1-41) subject to the constraint in (4.1-40) can be carried out analytically using the method of calculus of variations. Unfortunately, however, the resulting solution for $s_0(t)$ is in the form of an integral equation that does not lend itself to a

Table 4-1. Bandwidth-efficient performance of TCM with prescribed decoding delay.

Signal	$1/B_{99}T_b$ ([b/s]/Hz)	% Improvement over MSK	$1/B_{99.9}T_b$ ([b/s]/Hz)	% Improvement over MSK
MSK	0.845	—	0.366	—
Optimum ($\nu = 1$)	0.896	6.04	0.659	79.7
Optimum ($\nu = 2$)	1.23	45.6	—	—

[7] Of course, in reality the continuous time-limited signal would have infinite bandwidth.

[8] Note that the definition of bandwidth B in Ref. 1 is one-half of that used in this monograph.

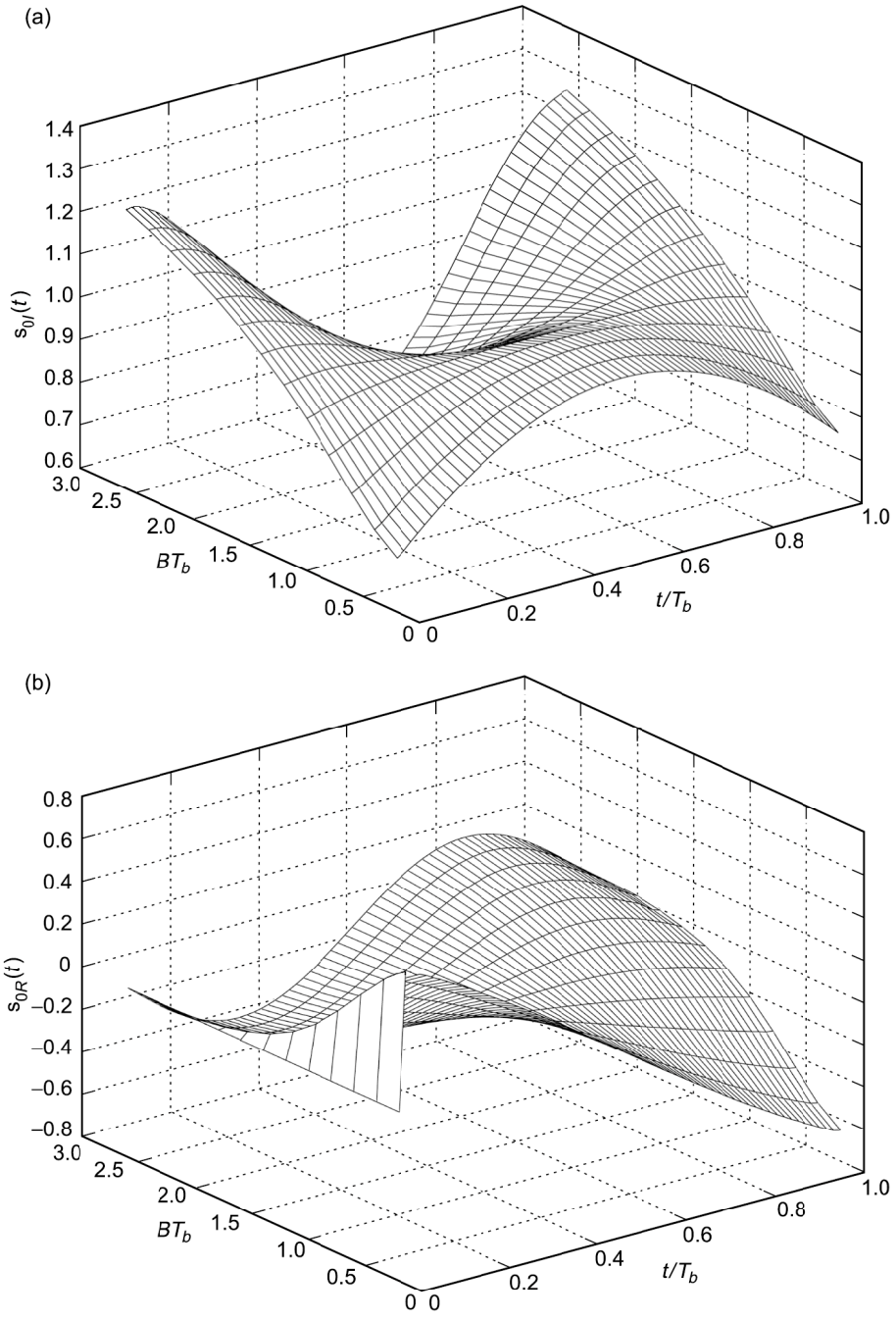

Fig. 4-3. Profiles of the optimum signal as a function of the bandwidth-bit time product: (a) the imaginary part and (b) the real part.

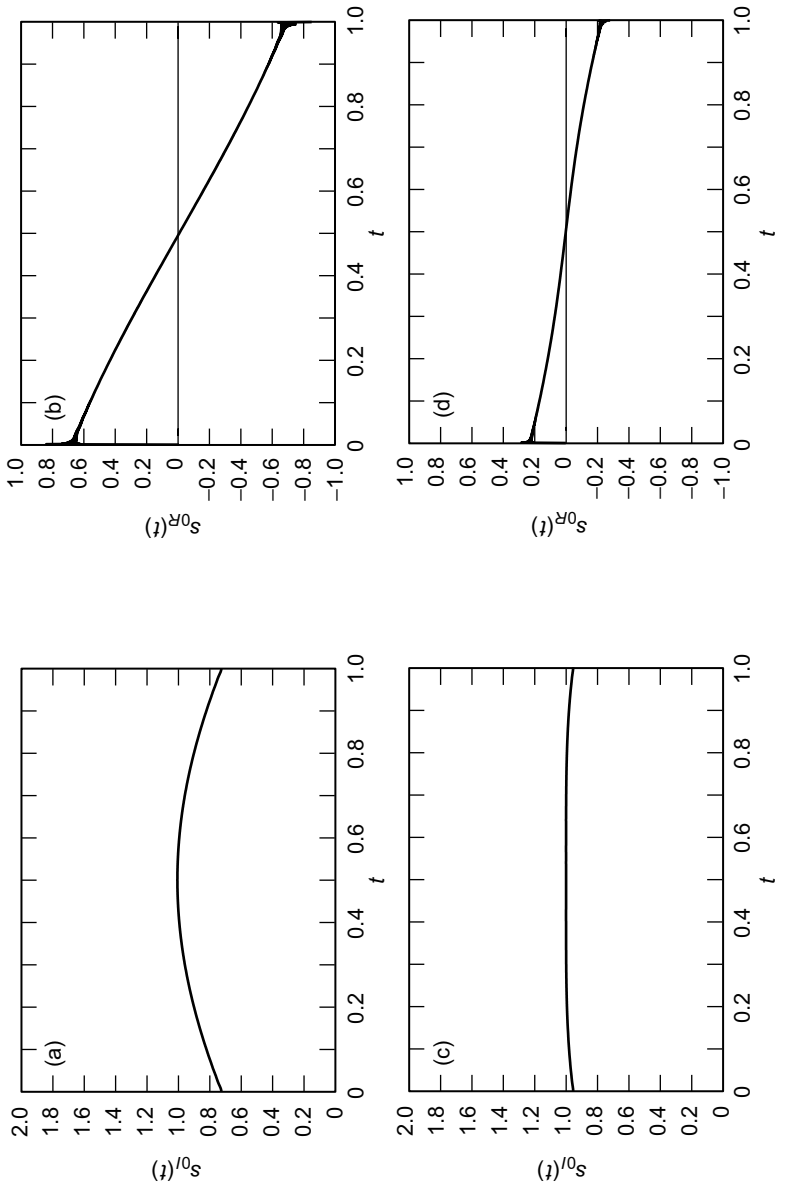

Fig. 4-4. The optimum signal for bandwidth-time product: (a) the imaginary part of the optimum signal for bandwidth-time product = 0.2, (b) the real part of the optimum signal for bandwidth-time product = 0.2, (c) the imaginary part of the optimum signal for bandwidth-time product = 1.0, and (d) the real part of the optimum signal for bandwidth-time product = 1.0.

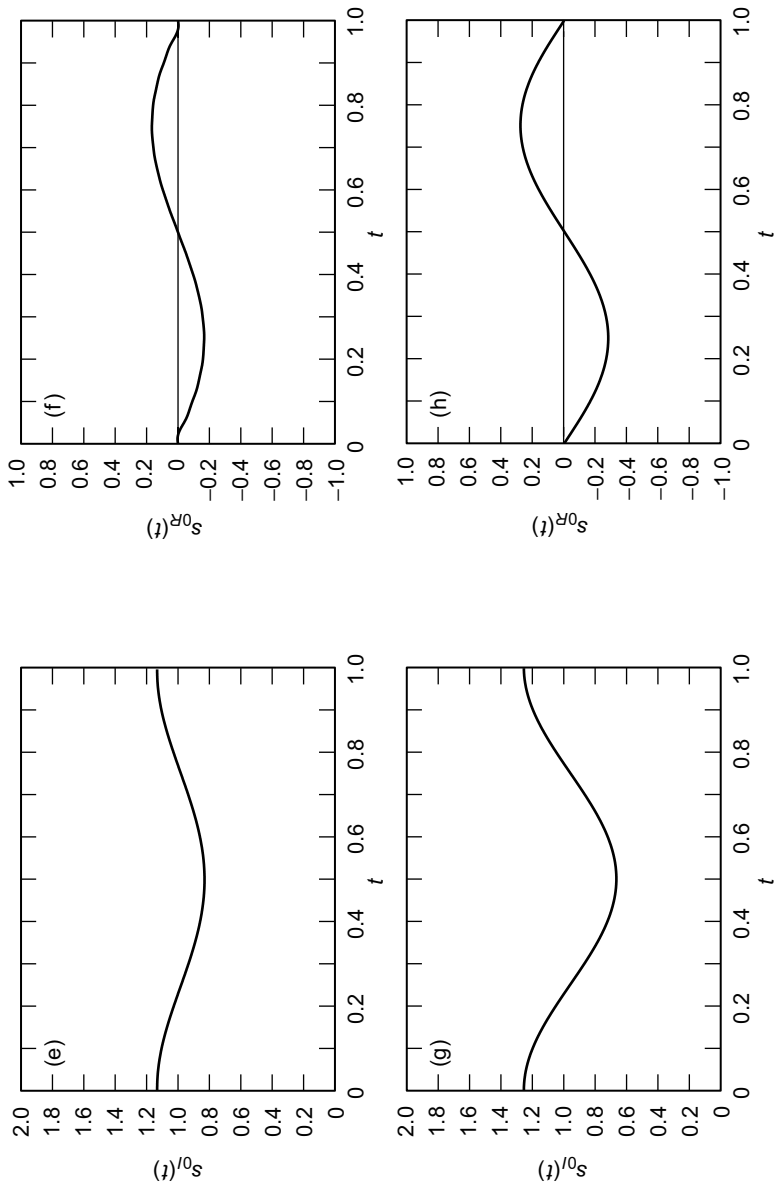

Fig. 4-4 (cont'd). The optimum signal for bandwidth-time product: (e) the imaginary part of the optimum signal for bandwidth-time product = 1.8, (f) the real part of the optimum signal for bandwidth-time product = 1.8, (g) the imaginary part of the optimum signal for bandwidth-time product = 2.6, and (h) the real part of the optimum signal for bandwidth-time product = 2.6.

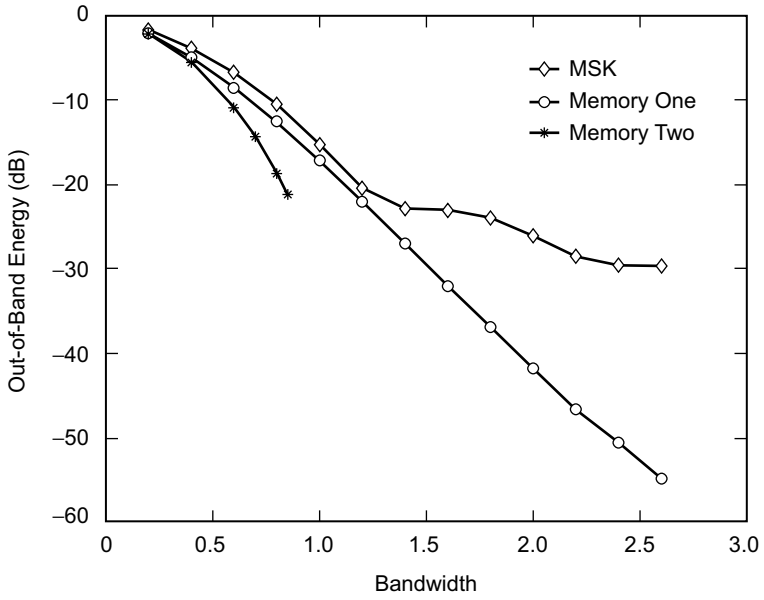

Fig. 4-5. Comparison of fractional out-of-band powers.

closed-form solution. Thus, there is no strong advantage to presenting these results here since we have already obtained a numerical solution as discussed above by direct maximization of (4.1-41). One interesting observation does result from applying the calculus of variations approach: $s_{0R}(t)$ is an odd function around its midpoint (at $t = 1/2$) and $s_{0I}(t)$ is an even function around its same midpoint. Clearly, this observation is justified by the numerical results illustrated in the various parts of Fig. 4-4.

4.1.3.2 Memory Two Case. Analogous to what was done for the memory one case, we need to maximize the fractional in-band power of (4.1-39), using now (4.1-31) for $G(f)$. Expressing the various Fourier transforms of (4.1-31) in terms of their associated signal waveforms and then performing the integration on frequency between $-B/2$ and $B/2$ as required in (4.1-39) produces the following result (again normalizing $T_b = 1$):

$$\int_{-B/2}^{B/2} G(f)\, df = \sum_{i=1}^{6} P_i \tag{4.1-43}$$

where

$$
\left.
\begin{aligned}
P_1 =& \frac{B}{4}\int_0^1\int_0^1 \left(s_1^{(2)}(t)+s_2^{(2)}(t)\right)\left(s_1^{(2)}(\tau)+s_2^{(2)}(\tau)\right)^* \\
&\times e^{-j(\pi/2)(t-\tau)}\text{sinc }\pi B(t-\tau)dtd\tau \\
P_2 =& \frac{B}{4}\int_0^1\int_0^1 \left(s_0^{(2)}(t)-s_2^{(2)}(t)\right)\left(s_0^{(2)}(\tau)-s_2^{(2)}(\tau)\right)^* \\
&\times e^{-j(\pi/2)(t-\tau)}\text{sinc }\pi B(t-\tau)dtd\tau \\
P_3 =& \frac{B}{4}\int_0^1\int_0^1 \left(s_0^{(2)}(t)-s_1^{(2)}(t)\right)\left(s_0^{(2)}(\tau)-s_1^{(2)}(\tau)\right)^* \\
&\times e^{-j(\pi/2)(t-\tau)}\text{sinc }\pi B(t-\tau)dtd\tau \\
P_4 =& 2\text{ Re}\left\{\frac{B}{4}\int_0^1\int_0^1 \left(s_0^{(2)}(t)-s_2^{(2)}(t)\right)\left(s_1^{(2)}(\tau)+s_2^{(2)}(\tau)\right)^* \right. \\
&\left.\times e^{-j(\pi/2)(t-\tau+1)}\text{sinc }\pi B(t-\tau+1)dtd\tau\right\} \\
P_5 =& 2\text{ Re}\left\{\frac{B}{4}\int_0^1\int_0^1 \left(s_0^{(2)}(t)-s_1^{(2)}(t)\right)\left(s_0^{(2)}(\tau)-s_2^{(2)}(\tau)\right)^* \right. \\
&\left.\times e^{-j(\pi/2)(t-\tau+1)}\text{sinc }\pi B(t-\tau+1)dtd\tau\right\} \\
P_6 =& 2\text{ Re}\left\{\frac{B}{4}\int_0^1\int_0^1 \left(s_0^{(2)}(t)-s_1^{(2)}(t)\right)\left(s_1^{(2)}(\tau)+s_2^{(2)}(\tau)\right)^* \right. \\
&\left.\times e^{-j(\pi/2)(t-\tau+2)}\text{sinc }\pi B(t-\tau+2)dtd\tau\right\}
\end{aligned}
\right\} \quad (4.1\text{-}44)
$$

From the constraint in (4.1-36b), $s_2^{(2)}(t)$ can be expressed in terms of $s_0^{(2)*}(t)$ and then substituted in (4.1-44). Thus, the optimization problem reduces to finding only two signals, $s_0^{(2)}(t)$ and $s_1^{(2)}(t)$, by joint maximization of (4.1-43) combined with (4.1-44). (Note that $s_3^{(2)}(t)$ can be found from $s_3^{(2)}(t) = s_2^{(2)}(t) - s_0^{(2)}(t) + s_1^{(2)}(t)$, once $s_0^{(2)}(t)$ and $s_1^{(2)}(t)$ are determined.)

Superimposed on Fig. 4-5 are the optimum fractional out-of-band power results for the memory two case. Due to the extremely time-consuming nature of the computer algorithms that perform the joint optimization procedure, particularly at low levels of fractional out-of-band power where extreme accuracy in satsifying the constraints is required, only results corresponding to values of $BT_b < 1$ (or equivalently $B < 1$ for $T_b = 1$) have been obtained thus far. Nevertheless, we are able to extract from these results the bandwidth-efficiency improvement relative to MSK for the 99 percent (-20 dB) out-of-band power level, and this improvement is included in Table 4-1. We observe that there is a significant improvement in out-of-band power performance, with no power efficiency penalty, by going from a memory one (1-bit decoding delay) modulation to one that has memory two (2-bit decoding delay).

4.2 Bandwidth-Efficient TCM with Prescribed Decoding Delay—Unequal Signal Energies

In the introduction to this chapter, we said that a relaxation of the equal energy condition on the signals could be used to potentially trade off between the power and bandwidth efficiency of the system. We now investigate the additional constraints that must be placed on the signals in order that the optimum TCM receiver still achieve a finite decoding delay equal to the memory of the modulation. In order to accomplish this, we first briefly review the received signal plus noise model, branch metric, and accompanying decision rule leading up to the conditions on the signal differences in Theorem I of Ref. 3 [summarized herein in (4.1-6) and (4.1-7)] and then modify them so as to apply to the case of unequal signal energies.

Corresponding to the baseband signal, $s(t)$, of (4.1-1) transmitted over an AWGN channel, the received signal is

$$R(t) = s(t) + N(t) \tag{4.2-1}$$

where $N(t)$ is again a zero-mean complex Gaussian noise process with PSD N_0 watts/hertz. For equal energy signals, the maximum-likelihood (Viterbi) receiver uses as its branch metric in the nth interval

$$\lambda_n(s_i) = \mathrm{Re}\left\{\int_{nT}^{(n+1)T} R^*(t)\, s_i(t - nT)\, dt\right\}$$

$$= \mathrm{Re}\left\{\int_0^T R^*(t + nT)\, s_i(t)\, dt\right\}, \quad i \in \{0, 1, \cdots, 2^{\nu+1} - 1\} \tag{4.2-2}$$

As previously stated, without any constraints on the signal set, for true optimality, the Viterbi receiver *theoretically* needs to observe the entire transmitted sequence (sum over an infinite number of branch metrics), resulting in an infinite decoding delay although in practice one may decode with finite delay using a truncated (but suboptimal) form of VA. If the signal differences are constrained as in (4.1-6) and (4.1-7), then, as previously stated in Theorem I of Ref. 3, the receiver can optimally decode the nth information symbol after ν symbol intervals, according to the decision rule:

$$\text{Choose } U_n = 0 \text{ if } \sum_{i=\nu}^{n+\nu} \lambda_i \left(s_0 - s_{2^{n+\nu-i}}\right) > 0, \quad \text{otherwise choose } U_n = 1 \quad (4.2\text{-}3)$$

For unequal energy signals, the branch metric of (4.2-2) would be modified to

$$\lambda_n (s_i) = \text{Re}\left\{ \int_{nT}^{(n+1)T} R^* (t)\, s_i (t - nT)\, dt \right\} - \frac{E_i}{2}$$

$$= \text{Re}\left\{ \int_0^T R^* (t + nT)\, s_i (t)\, dt \right\} - \frac{E_i}{2}, \quad i \in \left\{0, 1, \cdots, 2^{\nu+1} - 1\right\} \quad (4.2\text{-}4)$$

where $E_i = \int_0^T |s_i (t)|^2\, dt$ is the energy of the ith signal in the set. Since the derivation of the conditions for finite decoding delay given in Ref. 3 relies on comparisons of sums of branch metrics, it is straightforward to substitute (4.2-4) for (4.2-2) in the steps of this derivation, which leads to an additional set of conditions on the energies of the signals. To illustrate the procedure, we first consider the simplest case corresponding to unit memory ($\nu = 1$).

Consider the two-state trellis (corresponding to the nth and $n+1$st intervals) in Fig. 4-6, where each branch is labeled with: (a) the input bit that causes the transition between states and (b) the baseband signal transmitted in accordance with the choice defined in Fig. 4-1(b). Assume first that we are in state "0" at time n (having gotten there as a result of decoding symbols in the previous intervals). Suppose now that the two paths (of length two branches) that survive at time $n+2$ are those that merge at (eminate from) the same node at time $n+1$ (thereby allowing unique decoding of the transmitted symbol, U_n). Since this node can correspond to either state "0" or state "1," there exist two possibilities, which are indicated by heavy lines in Figs. 4-5(a) and 4-5(b).

For Fig. 4-6(a), both surviving paths have a first branch corresponding to $U_n = 1$ and, thus, the decision $\hat{U}_n = 1$ is unique provided that

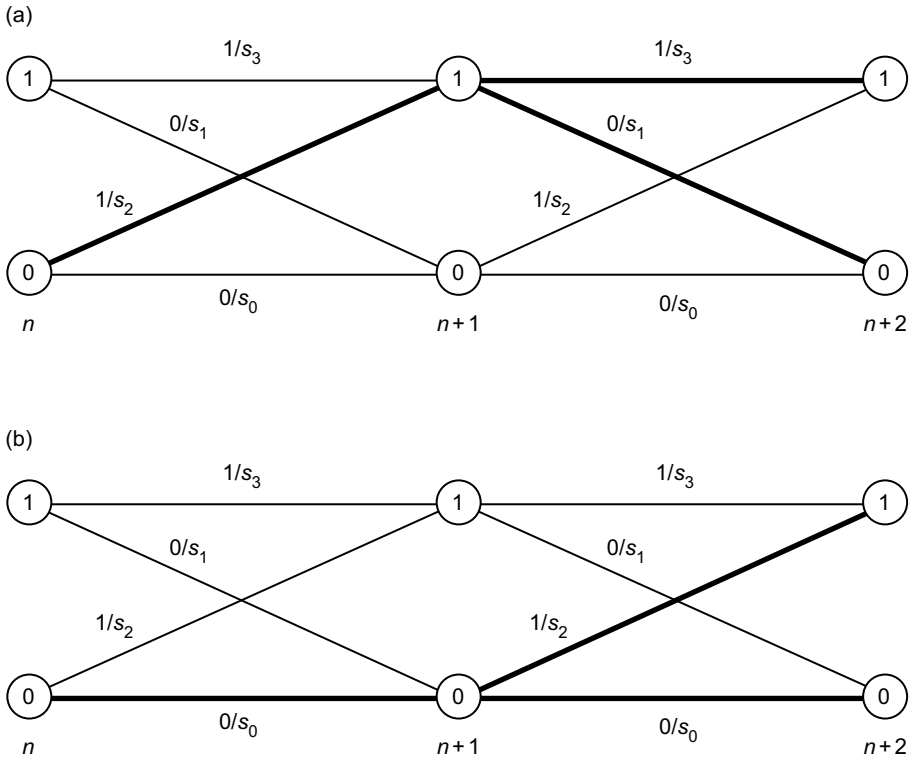

Fig. 4-6. A trellis diagram for memory one modulation, assuming state "0" at time *n*: (a) surviving paths merging at state "1" at time *n* + 1 and (b) surviving paths merging at state "0" at time *n* + 1.

$$\lambda_n(s_2) + \lambda_{n+1}(s_3) > \lambda_n(s_0) + \lambda_{n+1}(s_2) \tag{4.2-5a}$$

and

$$\lambda_n(s_2) + \lambda_{n+1}(s_1) > \lambda_n(s_0) + \lambda_{n+1}(s_0) \tag{4.2-5b}$$

or equivalently

$$\lambda_n(s_0) - \lambda_n(s_2) + \lambda_{n+1}(s_2) - \lambda_{n+1}(s_3) < 0 \tag{4.2-6a}$$

and

$$\lambda_n(s_0) - \lambda_n(s_2) + \lambda_{n+1}(s_0) - \lambda_{n+1}(s_1) < 0 \tag{4.2-6b}$$

To simultaneously satisfy (4.2-6a) and (4.2-6b), we need to have

$$\lambda_{n+1}(s_2) - \lambda_{n+1}(s_3) = \lambda_{n+1}(s_0) - \lambda_{n+1}(s_1) \tag{4.2-7}$$

which is the identical requirement found by Li and Rimoldi [3] when treating the equal signal energy case. Using instead now the metric definition in (4.2-4) for unequal energy signals, then analogous to the results in Ref. 3, the condition of (4.2-7) can be satisfied by the first equality in (4.1-5a), namely,

$$s_0(t) - s_1(t) = s_2(t) - s_3(t) \tag{4.2-8}$$

and, furthermore,

$$E_0 - E_1 = E_2 - E_3 \tag{4.2-9}$$

Note that the relation in (4.2-9) is identical in form to that in (4.2-8) if each of the signals in the latter is replaced by its energy. This observation will carry over when we consider modulations with memory greater than one.

For Fig. 4-6(b), both surviving paths have a first branch corresponding to $U_n = 0$ and thus the decision $\hat{U}_n = 0$ is unique provided that

$$\lambda_n(s_0) + \lambda_{n+1}(s_2) > \lambda_n(s_2) + \lambda_{n+1}(s_3) \tag{4.2-10a}$$

and

$$\lambda_n(s_0) + \lambda_{n+1}(s_0) > \lambda_n(s_2) + \lambda_{n+1}(s_1) \tag{4.2-10b}$$

or equivalently

$$\lambda_n(s_0) - \lambda_n(s_2) + \lambda_{n+1}(s_2) - \lambda_{n+1}(s_3) > 0 \tag{4.2-11a}$$

and

$$\lambda_n(s_0) - \lambda_n(s_2) + \lambda_{n+1}(s_0) - \lambda_{n+1}(s_1) > 0 \tag{4.2-11b}$$

It is clear that the condition in (4.2-7) will also simultaneously satisfy (4.2-11a) and (4.2-11b).

Finally, if we assume that we were in state "1" at time n, then it is straightforward to show that the conditions on the signal set that produce a unique decision on U_n would be identical to those in (4.2-8) and (4.2-9). Thus, we conclude that *for a memory one modulation of the type described by Fig. 4-1(b) with unequal energy signals, the conditions on the signal set to guarantee unique decodability with one symbol delay are those given in (4.2-8) and (4.2-9).*

To extend the above to modulations with memory ν greater than one, we proceed as follows: As was observed in Ref. 3, what we now seek are the inequality conditions on the sums of branch metrics such that the 2^ν surviving paths at time $n+\nu$ merge at a single node at time $n+1$. Given a particular state at time n, this set of 2^ν conditions then allows for uniquely decoding U_n. Since these conditions are expressed entirely in terms of the branch metrics for the surviving paths, and, as such, do not depend on the form of the metric itself (i.e., whether it be (4.2-2) for equal energy signals or (4.2-4) for unequal energy signals), then it is straightforward to conclude that the finite decoding delay conditions on the signal set derived in Ref. 3 for the equal energy case also apply now to the signal energies in the nonequal energy case. Specifically, in addition to (4.1-6), the signal set must satisfy the energy conditions

$$E_0 - E_{2^m} = E_{2^{m+1}l} - E_{2^{m+1}l+2^m},\ m = 0, 1, 2, \cdots, \nu - 1,\ l = 1, 2, \cdots, 2^{\nu-m} - 1 \tag{4.2-12}$$

For the equal energy case, (4.2-12) is trivially satisfied.

Having now specified the conditions for achieving finite decoding delay with unequal energy signals, we now investigate the impact of this relaxed restriction on the minimum-squared Euclidean distance (power efficiency) of the modulation. Again consider first the memory one case. For the trellis diagram of Fig. 4-5(a), the unnormalized squared Euclidean distance between the length 2 error event path and the all zeros path (corresponding to $U_n = 0, U_{n+1} = 0$) is

$$\begin{aligned} D^2 &= \int_0^T |s_0(t) - s_2(t)|^2\, dt + \int_0^T |s_0(t) - s_1(t)|^2\, dt \\ &= 2E_0 + E_1 + E_2 - 2\ \mathrm{Re}\left\{ \int_0^T s_0^*(t)\,(s_1(t) + s_2(t))\, dt \right\} \end{aligned} \tag{4.2-13}$$

Using (4.2-8) and (4.2-9) in (4.2-13) enables rewriting it in the form

$$\left.\begin{aligned} D^2 &= 2E_{av} - 2\ \mathrm{Re}\left\{\int_0^T s_0^*(t)\, s_3(t)\, dt\right\} \\ E_{av} &= \frac{E_0 + E_1 + E_2 + E_3}{4} = \frac{E_0 + E_3}{2} \end{aligned}\right\} \tag{4.2-14}$$

which when normalized by the average energy of the signal set, E_{av}, gives

$$d^2 \triangleq \frac{D^2}{2E_{av}} = 1 - \frac{\mathrm{Re}\left\{\int_0^T s_0^*(t)\, s_3(t)\, dt\right\}}{E_{av}} = 1 - \frac{\mathrm{Re}\left\{\int_0^T s_0^*(t)\, s_3(t)\, dt\right\}}{(E_0 + E_3)/2} \tag{4.2-15}$$

Following steps analogous to (4.2-13)–(4.2-15) and using the signal difference property in (4.2-8), it is straightforward to show that the unnormalized squared Euclidean distance between any pair of length 2 paths beginning and ending at the same node (i.e., other pairwise error events) is given by (4.2-15), i.e., the trellis has a uniform error probability (UEP) property. It can also be shown using a combination of (4.2-8) and (4.2-9) in (4.2-13) that (4.2-15) can be expressed as

$$d^2 \triangleq \frac{D^2}{2E_{av}} = 1 - \frac{\mathrm{Re}\left\{\int_0^T s_1^*(t)\, s_2(t)\, dt\right\}}{(E_1 + E_2)/2} \tag{4.2-16}$$

Finally noting that $-1 \le \mathrm{Re}\left\{\int_0^T s_0^*(t)\, s_3(t)\, dt\right\} / \left[(E_0 + E_3)/2\right]$ with equality achieved when $s_0(t) = -s_3(t)$ and, likewise, $-1 \le \mathrm{Re}\left\{\int_0^T s_1^*(t)\, s_2(t)\, dt\right\} / \left[(E_1 + E_2)/2\right]$ with equality achieved when $s_1(t) = -s_2(t)$, then, in order to achieve the maximum value, $d^2_{\min} = 2$, we would need to choose $s_0(t) = -s_3(t)$, which produces $E_0 = E_3$ and also $s_1(t) = -s_2(t)$, which produces $E_1 = E_2$. However, from (4.2-9), $E_0 + E_3 = E_1 + E_2$ and, thus, $E_0 = E_1 = E_2 = E_3 = E$, i.e., all signals have equal energy. Therefore, we conclude that *for memory one, an unequal energy signal set necessarily results in a value of* $d^2_{\min} < 2$.

For arbitrary memory, ν, by a straightforward extension of the procedure for memory one, it can be shown that the distance between any pair of length $\nu + 1$ paths beginning and ending at the same node (i.e., pairwise error events) is, analogous to (4.2-12), given by

$$d^2 \triangleq 1 - \frac{\mathrm{Re}\left\{\int_0^T s_0^*(t)\, s_{2^{\nu+1}-1}(t)\, dt\right\}}{(E_0 + E_{2^{\nu+1}-1})/2} \tag{4.2-17}$$

Thus, to achieve the maximum value, $d^2_{\min} = 2$, we would need to choose $s_0(t) = -s_{2^{\nu+1}-1}(t)$, which produces $E_0 = E_{2^{\nu+1}-1}$. However, in view of the other forms [analogous to (4.2-16)] that (4.2-17) can be expressed as, it can also be shown that achieving $d^2_{\min} = 2$ also requires choosing $s_i(t) = -s_{2^{\nu+1}-1-i}(t), i = 1, 2, \cdots, 2^\nu - 1$, which produces $E_i = E_{2^{\nu+1}-1-i}, i = 1, 2, \cdots, 2^\nu - 1$. Finally, using the energy conditions in (4.2-12), we arrive at the fact that $d^2_{\min} = 2$ can only be achieved when $E_0 = E_1 = E_2 = \cdots = E_{2^{\nu+1}-1} = E$, i.e., all signals have equal energy. Thus, we conclude that *for arbitrary memory, an unequal energy signal set necessarily results in a value of* $d^2_{\min} < 2$.

References

[1] A. J. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. IT-13, no. 2, pp. 260–269, April 1967.

[2] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*, New York: McGraw-Hill, Inc., 1979.

[3] Q. Li and B. E. Rimoldi, "Bandwidth-efficient trellis-coded modulation scheme with prescribed decoding delay," *International Symposium on Information Theory*, Ulm, Germany, June 29–July 4, 1997.

[4] G. D. Forney, Jr., "The Viterbi Algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, March 1973.

[5] B. E. Rimoldi, "A decomposition approach to CPM," *IEEE Transactions on Information Theory*, vol. IT-34, no. 3, pp. 260–270, May 1988.

[6] M. K. Simon, S. M. Hinedi, and W. C. Lindsey, *Digital Communication Techniques: Signal Design and Detection*, Upper Saddle River, New Jersey: Prentice Hall, 1995.

[7] M. K. Simon, "A generalization of MSK-type signaling based upon input data symbol pulse shaping," *IEEE Transactions on Communications*, vol. COM-24, no. 8, pp. 845–856, August l976.

[8] F. Amoroso, "Pulse and spectrum manipulation in the minimum (frequency) shift keying (MSK) format," *EEE Transactions on Communications*, vol. COM-24, no. 3, pp. 381–384, March 1976.